EJMO

DOI: 10.14744/ejmo.2020.91768
EJMO 2020;4(4):336–348

Research Article

In Silico Identification of Potential Inhibitors of the Main Protease of SARS-CoV-2 Using Combined Ligand-Based and Structure-Based Drug Design Approach

Bimal Debnath,[1] Apu Kr. Saha,[2] Samhita Bhaumik,[3] Sudhan Debnath[4]

[1]Department of Forestry and Biodiversity, Tripura University, Suryamaninagar, Tripura, India
[2]Department of Mathematics, National Institute of Technology, Agartala, Tripura, India
[3]Department of Chemistry, Women's College, Agartala, Tripura, India
[4]Department of Chemistry, MBB College, Agartala, Tripura, India

Abstract

Objectives: The outbreak of coronavirus disease 2019 (COVID-19) caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) remains a serious global threat. At the time of writing, there are no specific therapeutic agents or vaccines to combat this disease. This study was designed to identify the SARS-CoV-2 main protease inhibitors using drug molecule information retrieved from DrugBank 5.0 (Wishart et al.)

Methods: A set of common pharmacophores were generated from a series of 22 known SARS-CoV inhibitors. The best pharmacophore used for virtual screening (VS) of DrugBank using the Phase module followed by structure-based virtual screening (VS) using Glide (Release 2020-1; Schrödinger LLC, New York, NY, USA) with SARS-CoV-2 main protease and 50 ns molecular dynamics (MD) simulation studies.

Results: Six hits were selected based on the fitness score, extra-precision Glide score, and binding affinity with the main protease (Mpro). The predicted inhibitor constant (Ki) values of the 3 best hits, DB03777, DB06834, and DB07456, were 0.8176, 0.2148, and 0.1006 μM, respectively. An MD simulation of DB07456 and DB13592 with the Mpro demonstrated stable protein-ligand complexes.

Conclusion: The selected inhibitors displayed a similar type of binding interaction with co-ligands and remdesivir, and the predicted Ki values of 2 inhibitors were found to be superior to remdesivir. These selected hits may be used for further in vitro and in vivo studies against the SARS-CoV-2 Mpro.

Keywords: COVID-19, DrugBank, molecular docking, molecular dynamics, pharmacophore, SARS-CoV-2, virtual screening

Cite This Article: *Debnath B, Saha AK, Bhaumik S, Debnath S. In Silico Identification of Potential Inhibitors of the Main Protease of SARS-CoV-2 Using Combined Ligand-Based and Structure-Based Drug Design Approach. EJMO 2020;4(4):336–348.*

The recent eruption and ensuing pandemic of coronavirus disease 2019 (COVID-19) caused by the new severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has presented unprecedented global public health and economic challenges. The disease was first detected in Wuhan, China, but has been identified in 212 countries and territories with 3,566,210 confirmed infected cases and 248,285 deaths worldwide as of May, 4, 2020. The World Health Organization (WHO) has estimated that the mortality rate of COVID-19 is 3.4%. In recent history, there have been 2 outbreaks of other coronavirus diseases, the SARS-CoV epidemic of 2003 that led to 1000 deaths[1] and the Middle Eastern respiratory syndrome (MERS-CoV) outbreak of 2012, which claimed 862 lives.[2] The WHO declared the current worldwide COVID-19 outbreak an international public health emergency. At the time of writing, it continues to spread beyond control. The SARS-CoV-2 is highly homologous to SARS-CoV and the fatality rate of SARS-CoV was 10%.[3, 4] SARS-CoV-2 is a single-strand positive-sense RNA virus of the Coronavirinae subfamily, Coronaviridae

Address for correspondence: Sudhan Debnath, PhD. Department of Chemistry, MBB College, Agartala, Tripura (W), 799004, India
Phone: +91-381-2516728 **E-mail:** bcsdebnath@gmail.com
Submitted Date: August 12, 2020 **Accepted Date:** September 24, 2020 **Available Online Date:** December 25, 2020

family, Nidovirales order. Extensive research of human coronaviruses has been performed by numerous researchers, who have identified 4 Coronavirinae genera (α, β, γ, and δ).[5, 6] The human coronaviruses HCoV-HKU1, HCoV–OC43, MERS-CoV, SARS-CoV, and SARS-CoV-2 are members of the betacoronavirus genus.[7] A detailed classification of human coronavirus is illustrated in Figure S1. The new SARS-CoV-2 has a 79.7% and 91.02% genomic similarity with SARS-CoV and the Malayan pangolin coronavirus (SARS-CoV-2 like), respectively.[8] It also has 96% and 89.6% identity with the envelope and nucleocapsid protein of SARS-CoV, respectively.[7] Similarities between SARS-CoV and SARS-CoV-2, such as the replication mechanism, may be useful to inhibit the SARS-CoV-2 infection cycle.

Potential antiviral therapies can be grouped in 2 categories, namely improving the human immune system, in which interferon plays a vital function, and targeting virus replication through blocking the signal pathways. Potential alternatives to control or prevent COVID-19 infection include vaccines, interferon therapies, small molecule drugs, peptide or oligonucleotide-based therapies, and monoclonal antibodies, as shown in Figure S2. Of several options to prevent COVID-19 disease, one of the important is the use of small-molecule drugs.[9, 10] In the present exigent situation, given that there is no specific drug or vaccine yet approved for COVID-19, existing drug molecules are being analyzed using systematic bioinformatic tools as a potential means to block the replication cycle of SARS-CoV-2. Researchers have examined several existing drug molecules and natural compounds on the basis of structure-based virtual screening (VS) of various databanks.[5, 9-14] In the present study, SARS-CoV-2 inhibitors were explored using resources from DrugBank (5.0; Wishart et al.) using a ligand-based and a structure-based VS virtual screening approach. For the ligand-based analysis, 22 SARS-CoV inhibitors were used to generate common pharmacophores for virtual screening.

The SARS-CoV genome contains 2 open reading frames (ORF1a and ORF1b) of viral replicase genes which encode 2 large, overlapping polyproteins, pp1a and pp1ab. These polyproteins produce a number of nonstructural proteins (Nsps) through proteolytic cleavage. The SARS-CoV encodes 2 proteases: chymotrypsin-like protease (3CLpro) and papain-like protease (PLpro). 3CLpro is also the main protease (Mpro); it is cleaved automatically from polyproteins and then regulates the downstream proteolytic cleavage reaction at 11 polyprotein sites and releases Nsp4 to Nsp16, which are crucial for virus replication and pathogenesis.[7, 5, 15, 16] Liu et al.[9] suggested that the Mpro is a potential drug target to inhibit SARS-CoV-2 replication due to its highly conserved sequence and available 3-dimensional (3D) structure.

The structural organizations of the SARS-CoV genomes are displayed in Figure S3. The Mpro is a key protein for the proteolytic maturation of virus Nsps. The chymotrypsin-like protease of SARS-CoV-2 was selected as the target protein for the structure-based method VS. Various 3CLpro inhibitors of SARS-CoV-2 have been reported by several research groups to be potentially effective treatment options for COVID-19 based on a structure-based approach.[17–20] The present work used the best 3CLpro hits from DrugBank for in vitro study.

Methods

Collection of Materials

The high resolution (2.16 Å) X-ray crystal structure of the SARS-CoV-2 Mpro or 3CLpro (PDB ID: 6LU7), consisting of 312 amino acid residues complexed with a peptide-like inhibitor (N3), was retrieved from the Research Collaboratory for Structural Bioinformatics Protein DataBank.[21] The initial genomic sequences analyses of SARS-CoV-2 indicate that it has a high level of sequence similarity with SARS-CoV. In all, 22 known SARS-CoV inhibitors were collected based on the literature[4] and used to build a common pharmacophore. The structures of the diverse SARS-CoV inhibitors are shown in Figure 1. The DrugBank database was used

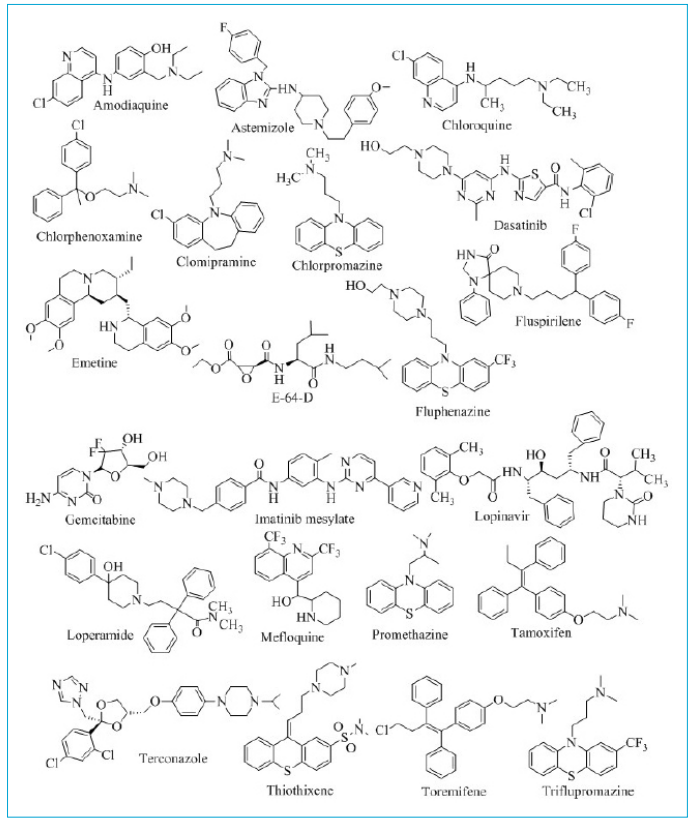

Figure 1. The 22 structurally diverse severe acute respiratory syndrome coronavirus 2 inhibitors used to generate common pharmacophore hypotheses.

for the ligand-based and structure-based VS.[22] The Mpro inhibitor remdesivir was also retrieved in the protein data bank (pdb) file format to compare with selected inhibitors. Windows 10 (Microsoft Corp., Redmond, WA, USA) and a 64-bit, Core 2 Duo CPU (Intel Corp., Santa Clara, CA, USA) microprocessor were used to perform the computations.

Ligand Preparation

All of the structures of the SARS-CoV inhibitors were drawn using ChemDraw Professional 15.1 (PerkinElmer Inc., Waltham, MA, USA) and saved in structure-data file (sdf) format. These inhibitors and DrugBank molecules were geometrically refined using the Schrödinger Ligprep module (Release 2020; Schrödinger LLC, New York, NY, USA).[23] During ligand preparation, the 2D structures were converted to the corresponding 3D structure, and added hydrogens and energy of ligands were minimized using the OPLS_2003 force field (Schrödinger LLC, New York, NY, USA) until it reached a root mean square deviation (RMSD) of 0.01 Å. Ligprep generates a single isomer with the original chiralities of each input. In order to perform molecular docking study using AutoDock 4.2 software (The Scripps Research Institute, La Jolla, CA, USA)[35] the selected molecules were converted to pdb format and then loaded in AutoDockTools 1.5.6 for conversion to pdbqt format (pdb with partial charges and AutoDock atom types).

Protein Preparation, Grid Generation and Docking Validation

The crystal structure of the Mpro (PDB ID: 6LU7) was used to prepare the protein using Protein Preparation Wizard (Schrödinger LLC, New York, NY, USA).[24] The protein structure was processed by assigning the bond orders, adding hydrogen, creating zero-order bonds to metals, filling side chains, and adding missing loops using the Prime protein structure prediction program (Schrödinger LLC, New York, NY, USA).[25-27] Finally, the protein was minimized using heavy atoms to RMSD 0.30 Å with OPLS3 force field. The receptor grid was prepared by selecting the coligand ligand of length 15Å from the center of the coligand. The coligand was split from its crystal structure and then docked with the prepared receptor grid. The best docked pose of the coligand was then superimposed on the original crystallographic bound conformation Figure S6 and the RMSD was calculated. The measured RMSD value was 2.0147Å. An RMSD value of <3Å is considered acceptable; therefore, the reproducibility of the docking protocol was good. In order to perform the molecular docking study using AutoDock 4.2, the protein was loaded in AutoDock Tools 1.5.6 and water molecules were removed, polar hydrogens bonded to heteroatoms were added, and the result was saved in pdbqt format.

Generation of Common Pharmacophore Hypothesis

The Phase[28-30] module (Release 2020-1; Schrödinger LLC, New York, NY, USA) was used to create common pharmacophores. Phase contains 6 in-built pharmacophore features: hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic group (H), negatively charged group (N), positively charged group (P), and aromatic ring (R). The 22 prepared SARS-CoV inhibitors were imported and common pharmacophore hypotheses were created by selecting multiple active ligands and using a minimum (Table 1) of 4 pharmacophore features. The best pharmacophore hypotheses are shown in Figure 2.

Ligand Based Virtual Screening

The best pharmacophore, HHRR.1, was used to search DrugBank[22] with VS to retrieve SARS–CoV pharmacophore-matched inhibitors. The VS workflow used to identify SARS–CoV pharmacophore-matched inhibitors is presented in Figure 3. The database compounds matching at

Table 1. Generated pharmacophoe hypothesis from 22 SARS-CoV inhibitors

HypoID	Survival	Site	Vector	Volume	Select	Matches
HHRR_1	4.6674	0.5523	0.8589	0.5774	1.5326	15
HHRR_2	4.6390	0.5503	0.8555	0.5665	1.5527	13
HHRR_3	4.4058	0.4738	0.798	0.4202	1.5377	14
HHRR_4	4.3177	0.5575	0.6887	0.4032	1.5221	14
HHRR_5	4.2837	0.4788	0.7121	0.3945	1.5222	15
HHRR_6	4.2764	0.5704	0.6129	0.4088	1.5381	14

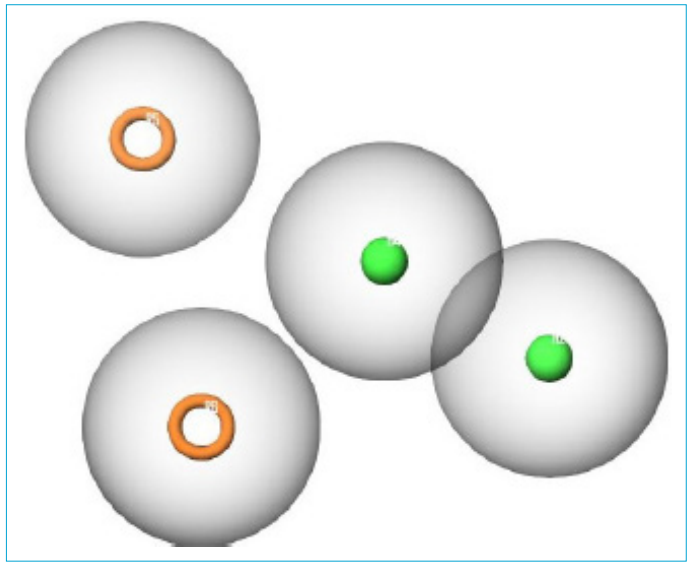

Figure 2. The Phase-generated best pharmacophore hypothesis, HHRR.1, illustrating hydrophobic (H2, H4: green) and aromatic rings (R5, R6: orange). In all, 22 known severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) inhibitors were used for this model.

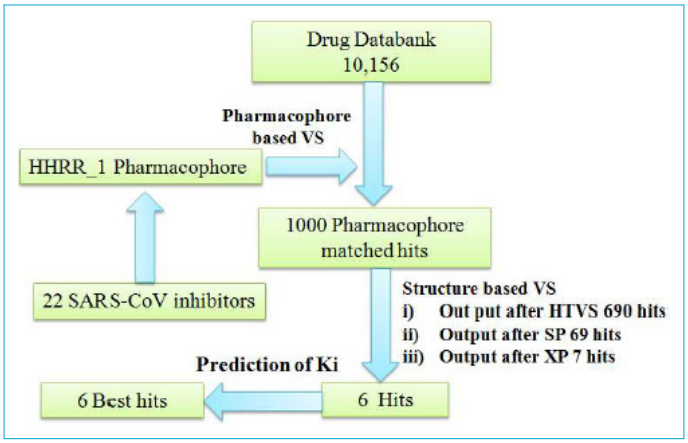

Figure 3. The virtual screening workflow to identify SARS-CoV-2 inhibitors. HTVS: High-throughput virtual screening; SP: Standard precision; VS: Virtual screening; XP: Extra precision.

least 4 pharmacophoric features were identified and 1000 hits were kept in reserve for structure-based VS.

Structure Based Virtual Screening

The output pharmacophore-matched hits were utilized for VS with Glide.[31-33] The input hits were filtered with QikProp (Schrödinger LLC, New York, NY, USA)[34] and the Lipinski rule, and reactive functionalities were removed. The filtered hits were used as an input hits for high-throughput VS and 10% of the best output hits were subjected to standard precision (SP) docking. Finally, 10% of the best SP outputs were used for extra precision (XP) docking. The VS workflow of structure-based VS is shown in Figure 3.

Molecular Dynamics Simulation (MD)

Molecular dynamics (MD) simulations to predict the stability of the protein-ligand complex and interaction analysis were performed using Desmond (Schrödinger LLC, D.E. Shaw Research, NY, NY, USA).[35] The best 3 ligand- Mpro complexes were placed in the orthorhombic box with a buffer distance of 10 Å in order to add water and TIP3P was used to generate water models.[36] The temperature and pressure of the system was kept at 300.15 K and 1.01325 bar, respectively. The cut-off for the van der Waals radius and electrostatic interactions was 9.0 Å and the time step was 2.0 fs. The 50 ns root-mean-square fluctuation (RMSF) MD simulations for the 3 complexes were performed with the isothermal-isobaric (NPT) ensemble using OPLS3e force field. The RMSD, RMSF, and interaction diagram were generated using the Maestro 12 interface (Schrödinger LLC, New York, NY, USA). The methods of simulation were conducted as described by Yoshino et al.[37]

Prediction of Ki Values of Selected Hits

The receptor was again prepared using AutoDock Tools 1.5.6 for docking with AutoDock 4.2. During the preparation, the water molecules were removed, and polar hydrogens were added, followed by computation of the Gasteiger charge and adding the Kollman charge. The best hits identified on the basis of the fitness score and XP Glide score were used to predict Ki values using AutoDock 4.2.[38]

Results and Discussion

A set of 6 pharmacophore hypotheses; HHRR_1, HHRR_2, HHRR_3, HHRR_4, HHRR_5, and HHRR_6, were generated from 22 SARS-CoV inhibitors. The survival score of HHRR_1 was the highest, 4.6674, which matched 15 SARS-CoV known inhibitors nicely as shown in Figure S4. The finding that of 22 SARS-CoV inhibitors, 15 fit with the best pharmacophore, indicates the quality of the model. DrugBank contains 10,156 compounds and the best 1000 pharmacophore-matched hits were subjected to structure-based VS. During VS, the hits were filtered using QikProp and the Lipinski rule, and reactive functional group compounds were removed. The best hits: DB06829, DB07456, DB13592, DB06834, DB11903, and DB03777, were selected based on a high XP Glide score. The 2D structures of these hits are shown in Figure 4. The range of fitness score of the selected hits was 1.816-1.459, and the range of the XP Glide score was -8.254 to -6.608. The fitness score, XP Glide score, and interacting active site amino acid residues and distance are shown in Table 2.

The fitness score of DB06829 was 1.753 and its XP Glide score was the highest, -8.254. This hit interacted with 2 key interacting amino acid residues, HIE-41 (2.10 Å) and CYS-145 (2.37 Å). The other interacting amino acid residues were HIS-163 (1.91Å) and ARG-188 (2.05Å). The 2D protein-ligand interaction (A), 3D protein-ligand interaction (B), pharmacophore match (C), and position in the active site of DB06829 are shown in Figure 5a.

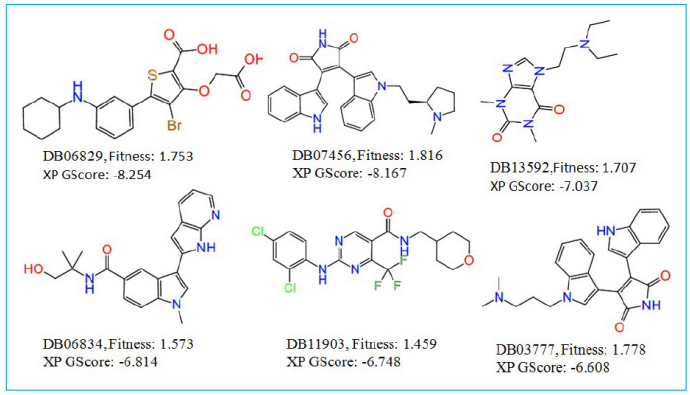

Figure 4. Selected severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) inhibitors with DrugBank ID, fitness with pharmacophore HHRR.1, and extra precision (XP) Glide score.

Table 2. Fitness score, XP Glide Score and different type of interactions

Drug Bank ID	Fitness XP	GScore	Interacting active site amino acid residues with interacting distance (Å)
DB06829	1.753	-8.254	HIE-41(π-π), GLY-143 (2.68), SER-144 (2.71), CYS-145 (2.58), GLU-166 (1.75), ARG-188 (2.03)
DB07456	1.816	-8.167	HIE-41 (two π-π), LEU-141 (2.35), CYS-145 (2.46)
DB13592	1.707	-7.037	HIE-41(2.36), SER-144 (2.23), CYS-145 (2.39)
DB06834	1.573	-6.814	HIE-41 (two, π-π), HIS-163 (2.03).
DB11903	1.459	-6.748	GLY-143 (2.19), GLN-189 (2.16)
DB03777	1.778	-6.608	HIE-41 (three π-π), SER-144 (2.42)
Remdesivir	-	-8.061	HIE-41 (2.13), HIS-164 (1.98), GLU-166 (2.47), THR-190 (1.98)

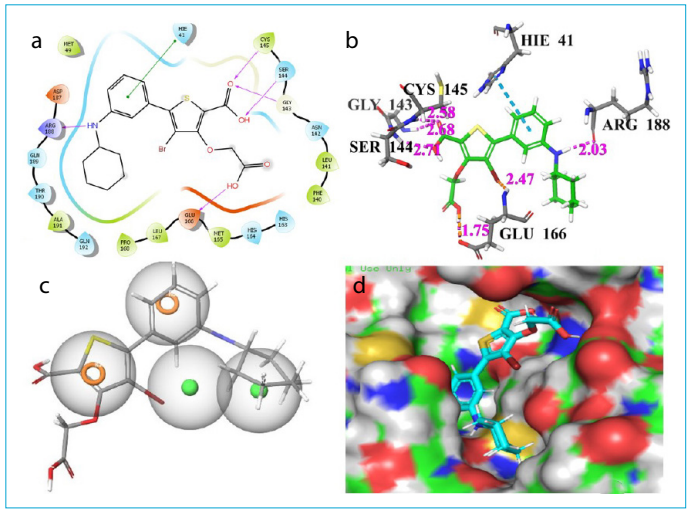

Figure 5a. Docking poses of compound DB06829 **(a)** 2D ligand interaction diagram of hydrogen bond donor-hydrogen bond acceptor (purple line) and π-π stacking (green line); **(b)** 3D ligand interactions of hydrogen bond donor-hydrogen bond acceptor (purple dotted line) and π-π stacking with distance (light blue dotted line); **(c)** Pharmacophore-match structure; **(d)** Position of selected hits in the active site.

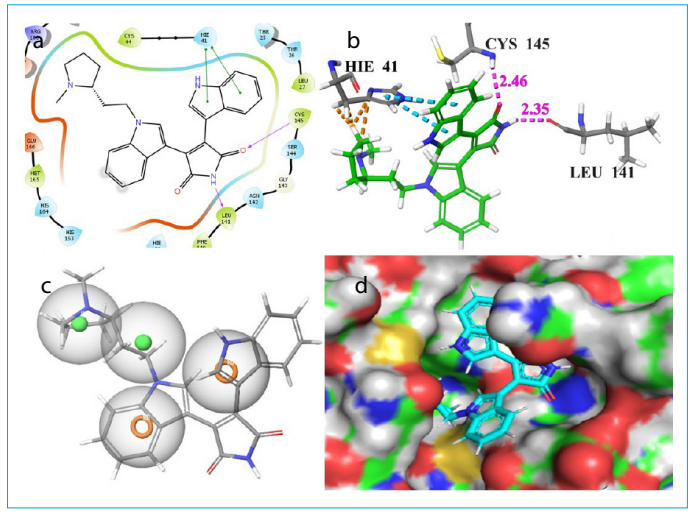

Figure 5b. Docking poses of compound DB07456 **(a)** 2D ligand interaction diagram of hydrogen bond donor-hydrogen bond acceptor (purple line) and π-π stacking (green line); **(b)** 3D ligand interactions of hydrogen bond donor-hydrogen bond acceptor (purple dotted line) and π-π stacking with distance (light blue dotted line); **(c)** Pharmacophore-match structure; **(d)** Position of selected hits in the active site.

The fitness score of DB07456 was 1.816 and the XP Glide score was the next highest at -8.167. This hit also interacted with 2 key amino acid residues of the SARS-CoV-2 Mpro, HIE-41 (2, π-π) and CYS-145 (2.46Å). The other interacting amino acid residue was LEU-141 (2.35Å). The 2D protein-ligand interaction (A), 3D protein-ligand interaction (B), pharmacophore match (C), and position in the active site of DB07456 are shown in Figure 5b.

The fitness score of DB13592 was 1.707 and the XP Glide score was -7.037. This hit interacted with HIE-41(2.36Å), SER-144 (2.23Å), and CYS-145 (2.39Å). The 2D protein-ligand interaction (A), 3D protein-ligand interaction (B), pharmacophore-match structure (C), and position in the active site of DB06834 are shown in Figure 5c.

The fitness score of compound DB06834 was 1.573 and its XP glide score was –6.814. This hit interacts with active site amino acid residues were HIE-41(two, π-π), HIS-163 (2.03Å). The 2D protein-ligand interaction (A), 3D protein-ligand in-

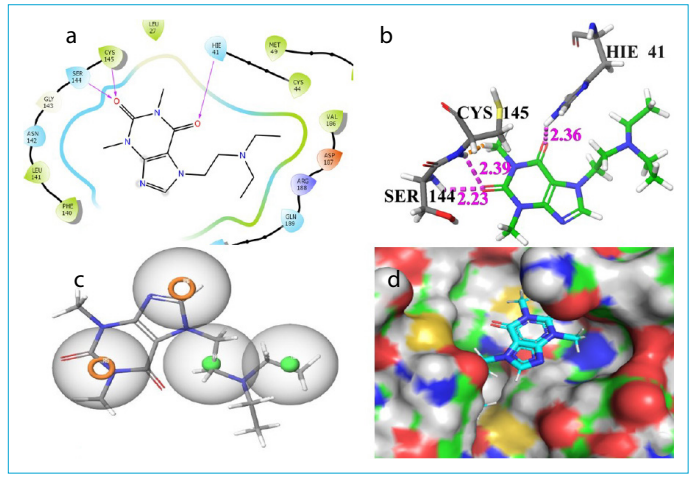

Figure 5c. Docking poses of compound DB13592A **(a)** 2D ligand interaction diagram of hydrogen bond donor-hydrogen bond acceptor (purple line) and π-π stacking (green line); **(b)** 3D ligand interactions of hydrogen bond donor-hydrogen bond acceptor (purple dotted line) and π-π stacking with distance (light blue dotted line); **(c)** Pharmacophore-match structure; **(d)** Position of selected hits in the active site.

teraction (B), pharmacophore-match structure (C) and position in the active site of DB06834 are shown in Figure 5d.

The fitness score of DB11903 was 1.459 and the XP Glide score was -6.748. This hit interacted with GLY-143 (2.19Å) and GLN-189 (2.16Å). The 2D protein-ligand interactions (A), 3D protein-ligand interaction (B), pharmacophore-match structure (C), and positions in the active site of DB11903 are shown in Figure 5e.

The fitness score of DB03777 was 1.778 and the XP Glide score was -6.608. This hit interacted with active site amino acids HIE-41 (3, π-π) and SER-144 (2.42 Å). The 2D protein-ligand interaction (A), 3D protein-ligand interaction (B), pharmacophore-matched structure (C), and position in the active site of DB03777 are shown in Figure 5f.

The coronavirus has more than a dozen proteins for viral entry and replication, including a papain-like protease (PLpro), a 3-chymotrypsin-like protease (3CLpro), an RNA-dependent RNA polymerase (RdRp), and a spike protein. SARS-3CLpro is a cysteine protease crucial to viral replication in the infection cycle.[39] Here, we targeted SARS-CoV-2 3CLpro (6LU7) to identify potential inhibitors. All 6 hits: DB06829, DB07456, DB07458, DB13592, DB06834, DB11903, and DB03777, had a common pharmacophore with SARS-CoV and good binding affinity with the active site of the Mpro. The active site of the 3CLpro has a Cys-His catalytic dyad (CYS-145 and HIS-41).[40] There was a covalent interaction between CYS-145 and the coligand (N3) in the original crystal structure of 6LU7 (Fig. S5). The interactions of remdesivir, an important SARS-CoV-2 ongoing clinical trial inhibitor,[41] were HIE-41, HIS-164, GLU-160, and THR-198.

Remdesivir also interacted with HIE-41, one of the important residues (Fig. S7). Therefore, HIE-41 and CYS-145 were crucial target amino acid residues for inhibition. Three selected hits, DB06829, DB07456, and DB13592, had close nonbonding interactions with CYS-145. The hits DB06829, DB07456, DB13592, DB06834, and DB03777 had close

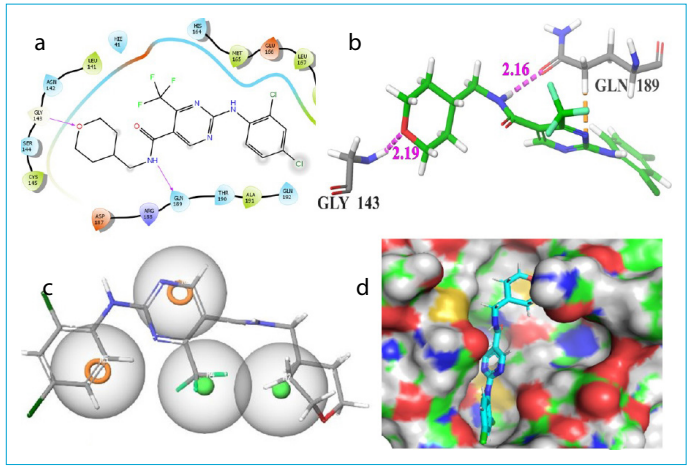

Figure 5e. Docking poses of compound DB11903 **(a)** 2D ligand interaction diagram of hydrogen bond donor-hydrogen bond acceptor (purple line) and π-π stacking (green line); **(b)** 3D ligand interactions of hydrogen bond donor-hydrogen bond acceptor (purple dotted line) and π-π stacking with distance (light blue dotted line); **(c)** Pharmacophore-match structure; **(d)** Position of selected hits in the active site

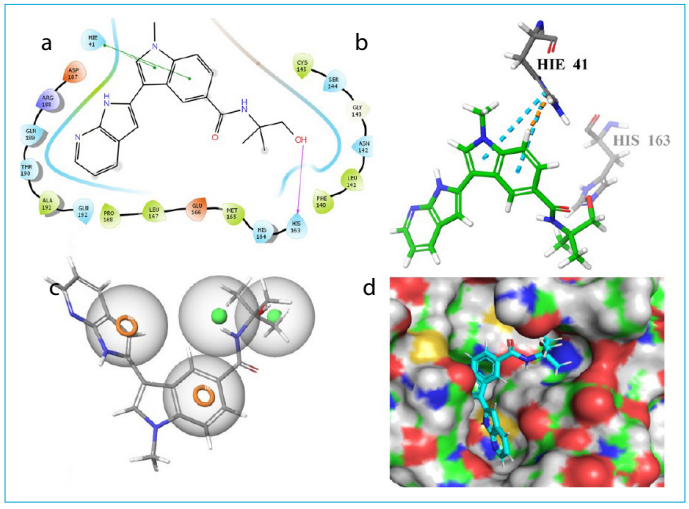

Figure 5d. Docking poses of compound DB06834 **(a)** 2D ligand interaction diagram of hydrogen bond donor-hydrogen bond acceptor (purple line) and π-π stacking (green line); **(b)** 3D ligand interactions of hydrogen bond donor-hydrogen bond acceptor (purple dotted line) and π-π stacking with distance (light blue dotted line); **(c)** Pharmacophore-match structure; **(d)** Position of selected hits in the active site.

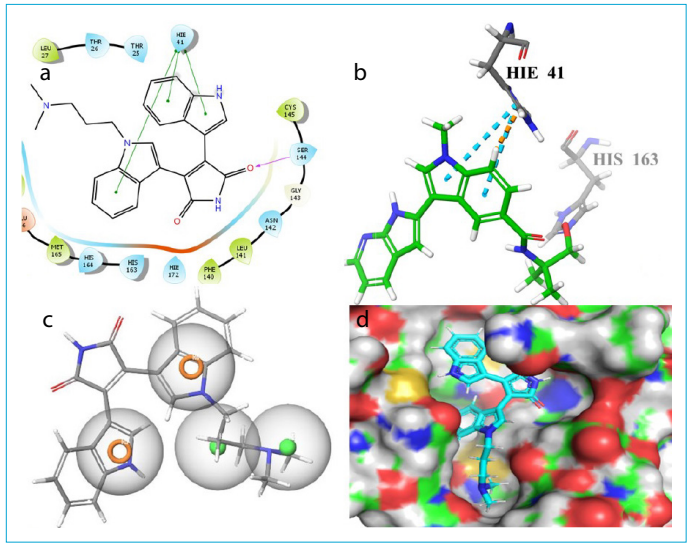

Figure 5f. Docking poses of compound DB03777A **(a)** 2D ligand interaction diagram of hydrogen bond donor-hydrogen bond acceptor (purple line) and π-π stacking (green line); **(b)** 3D ligand interactions of hydrogen bond donor-hydrogen bond acceptor (purple dotted line) and π-π stacking with distance (light blue dotted line); **(c)** Pharmacophore-match structure; **(d)** Position of selected hits in the active site

nonbonding interactions with HIS-41. The hits DB06829, DB07456, and DB13592 interacted with both CYS-145 and HIS-41. The hit DB06829 showed 1 π-π interaction and 3 closed, H-bond donor-acceptor interactions with an interacting distance of <3.0 Å. The hit DB07456 showed 2 π-π interactions, and 4 H-bond donor-acceptor interactions. Hit DB13592 demonstrated 3 H-bond donor-acceptor interactions from a close distance. The hits DB06834 and DB11903 revealed 2 H-bond donor-acceptor interactions at a distance of <3.0Å. All of the selected hits had a common pharmacophore with SARS-CoV, good binding affinity, and were well packed inside the active site pockets of the SARS-CoV-2 Mpro. The targets of all of the inhibitors are listed in Table S1.

The binding energy values of the selected hits B06829, DB07456, DB13592, DB06834, DB11903, DB03777, and remdesivir predicted by AutoDock were -8.22, -9.55, -6.04, -9.10, -7.89, -9.67, and -9.05, respectively. The predicted Ki values of B06829, DB07456, DB13592, DB06834, DB11903, DB03777, and remdesivir were 0.9400, 0.1006, 37.300, 0.2148, 1.6500, 0.0817, and 0.2334, respectively, as illustrated in Table 3. The binding energy and Ki values of DB07456, DB06834, and DB03777 were greater than those of remdesivir. DB07456 interacted with both CYS-145 and HIS-41; its XP Glide score was the second greatest, and its Ki value was 0.1006. Therefore, the hit DB07456 may be the best SARS-CoV-2 protease inhibitor. On the basis of fitness score, XP Glide score, predicted Ki values, interacting amino acid residues, and comparison with remdesivir, the 3 best hits, DB06829, DB07456, and DB13592, were selected for 50 ns MD simulation.

Overall information about the stability of the protein backbone after the formation of a protein-ligand complex can be analyzed according to the RMSD parameters. MD simulation of both apoprotein and the 3 best Mpro-ligand complex hits obtained from VS were monitored and analyzed for 50 ns to understand the dynamic behavior and stability of the protein-ligand complexes. The RMSD of the apoprotein and protein-ligand complexes of DB07456, DB06829, and DB13592 are shown in Figure 6. Important MD parameters, such as RMSD, RMSF, 2D-ligand-protein interaction diagram, and a timeline representation of interactions of the ligand with different active site amino acid residues, were calculated from the 50 ns MD trajectory. The protein-ligand complexes of DB07456, DB06829, and DB13592 exhibited an average RMSD value of 1.90 Å, 1.95 Å, and 1.68 Å, respectively. The RMSD of DB13592 was similar to the RMSD value of the apoprotein (1.68 Å). The average RMSD value of the ligand-Mpro complexes of compounds DB07456, DB06829, and DB13592 and the ligand-free protein are given in Table 4. The RMSD of the DB07456-Mpro system gradually increased for 7 ns, after which the value was lower than the RMSD of the apoprotein. At 16-25 ns, the RMSD value was slightly higher than the apoprotein RMSD. After that, the RMSD value was less than that of the apoprotein RMSD for the remainder of the simulation. The RMSD of the DB06829-Mpro system was similar to the apoprotein RMSD, but after 25 ns, the RMSD value was greater than the apoprotein RMSD. The RMSD of the DB13592-Mpro system was greater for the initial 15 ns, and then the value was very close to that of the apoprotein RMSD value. The lower value of RMSD of DB07456 after 15 ns and the RMSD value of DB13592 were very close to that of the apoprotein throughout the 50 ns MD simulation, which

Table 3. The binding energy and predicted Ki values of selected inhibitors using Autodock 4.2

Drug Bank ID	Binding Energy	Predicted Ki (μM)
DB06829	-8.22	0.9378
DB07456	-9.55	0.1006
DB13592	-6.04	37.300
DB06834	-9.1	0.2148
DB11903	-7.89	1.6500
DB03777	-9.67	0.0817
Remdesivir	-9.05	0.2334

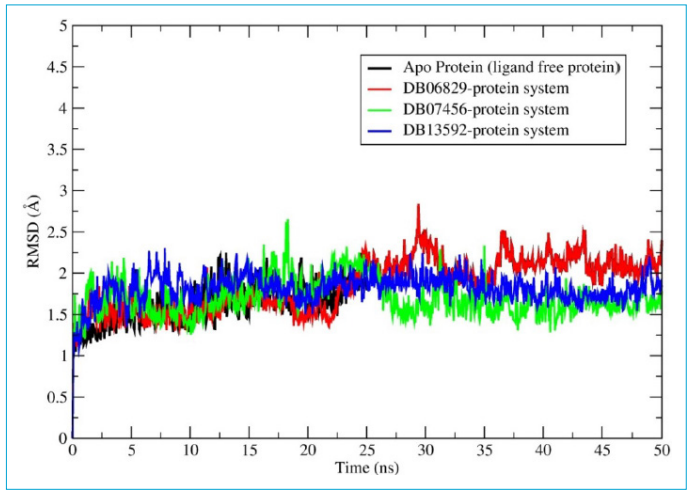

Figure 6. Root mean square deviation (RMSD) plot of apoprotein [main protease (Mpro), black], DB07456-Mpro complex (green), DB06829-Mpro complex (red), and DB13592-Mpro complex (blue) during 50 ns molecular dynamics simulation.

Table 4. The average RMSD and RMSF values of 50 ns MD simulations

Ligand No	Average RMSD (Å)	Average RMSF (Å)
Apo-protein (6lu7)	1.68	1.80
DB07456	1.90	1.25
DB06829	1.95	1.80
DB13592	1.68	1.35

indicates the high stability of the protein-ligand complexes. Therefore, the RMSD of the Mpro-DB07456 and Mpro-B13592 systems might indicate that they probably did not undergo significant conformational changes during the MD simulation. The acceptable range for an RMSD value is <3.0 Å;[42] therefore, all 3 protein-ligand complexes were within the acceptable range.

The average RMSF value of the apoprotein was 1.80 Å and the average RMSF value of DB07456 and DB13592 was 1.25Å and 1.35Å, respectively, which was less than that of the apoprotein. The RMSF value of DB06829 was 1.80 Å, which was similar to that of the apoprotein. The average RMSF values of the protein-ligand complexes of DB07456, DB13592, DB06829, and the apoprotein are shown in Figure 7 and Table 4. As seen in Figure 7, it is clear that there were no significant fluctuations of amino acid residues after binding in the active site. The greater RMSF values suggest that the protein structure is more flexible. The marginal flexibility of the protein-ligand system resulted in lower RMSF values.

The active site amino acid residues of the Mpro contributing to binding interactions with DB07456 were HIS-41 (π-π), THR-26 (hydrogen bond acceptor), ASN-142 (hydrogen bond acceptor), and GLU-166 (water bridge H-bond) produced 54%, 91%, 48%, and 44%, respectively. The interaction of DB06829 with HIS-41 yielded 44% and DB13592 with THR-190 resulted in 34%. The DB07456-Mpro, DB06829-Mpro, and DB13592-Mpro interaction patterns observed during the 50 ns MD simulations are shown in Figure S8. The contacts formed in the DB07456-Mpro complex over the course of the simulation were THR-26, HIS-41, ASN-142, and GLU-166. The contacts made by DB06829 with apoprotein the Mpro residues were THR-26, HIS-41, ASN-142, GLY-143, GLU-166, ARG-188, and THR-190. The DB13592 hit contacts with Mpro residues were HIS-41, MET-49, GLU-166, and GLN-189. All of the contacts formed by DB07456, DB06829, and DB13592 are presented in the top and bottom panels in Figure 8, illustrating how the residues interacted with the ligand in each trajectory frame during the simulation time. Wang[43] proposed that 1 hot spot residue, HIS-41, in the Mpro is a conserved residue across many viruses, including SARS-CoV, SARS-CoV-2, MERS-CoV, and hepatitis C virus (HCV). The selected hits DB07456 and DB06829 also showed fair interactions with HIS-41. It was observed that DB07456 and DB13592 had stable binding affinity throughout the MD simulation.

Conclusion

In the emergency situation of a COVID-19 pandemic, repurposing existing drugs may be helpful in the effort to

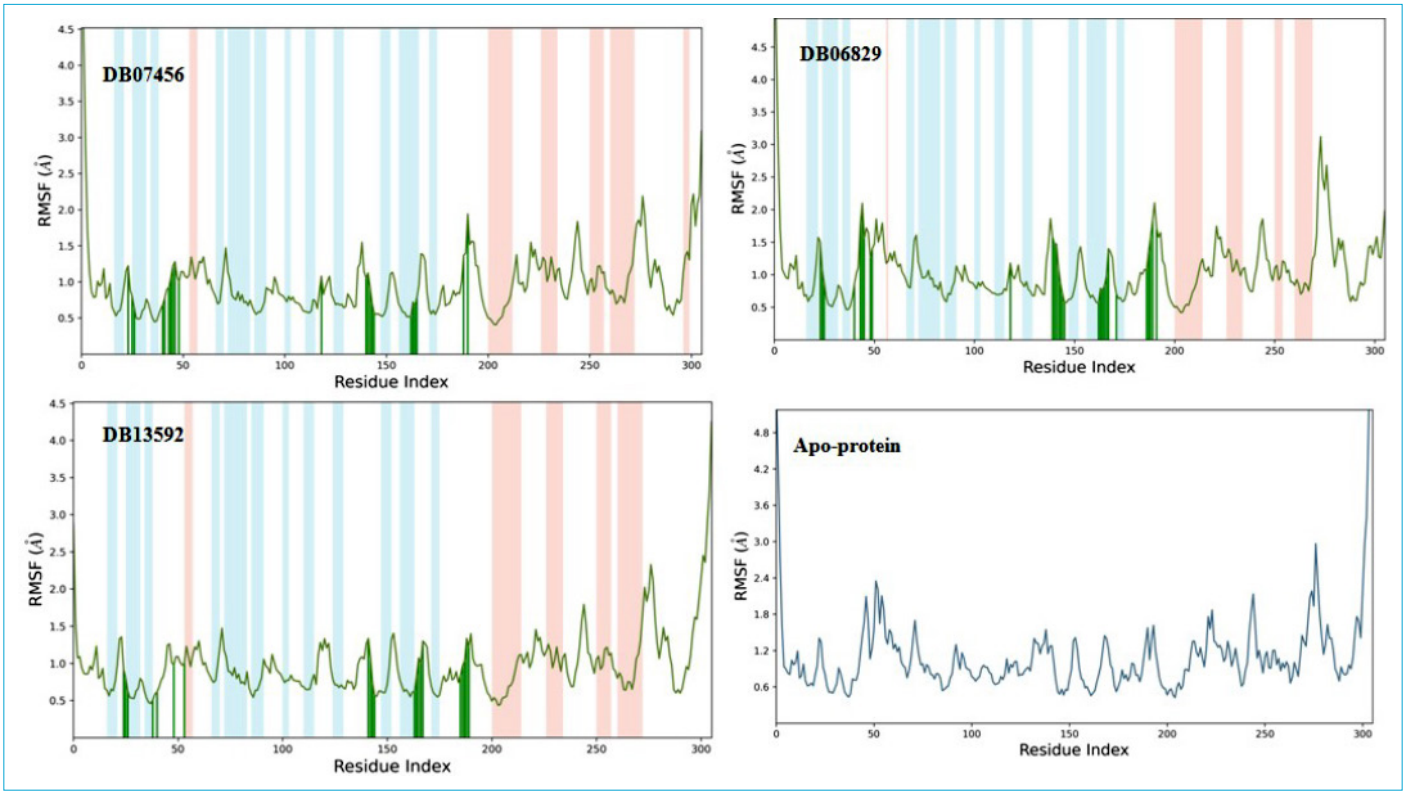

Figure 7. Root mean square fluctuation (RMSF) plot of apoprotein [main protease (Mpro)] residues, DB07456-Mpro complex, DB06829-Mpro complex, and DB13592-Mpro complex. Green line represents the binding site interaction with compound 16.

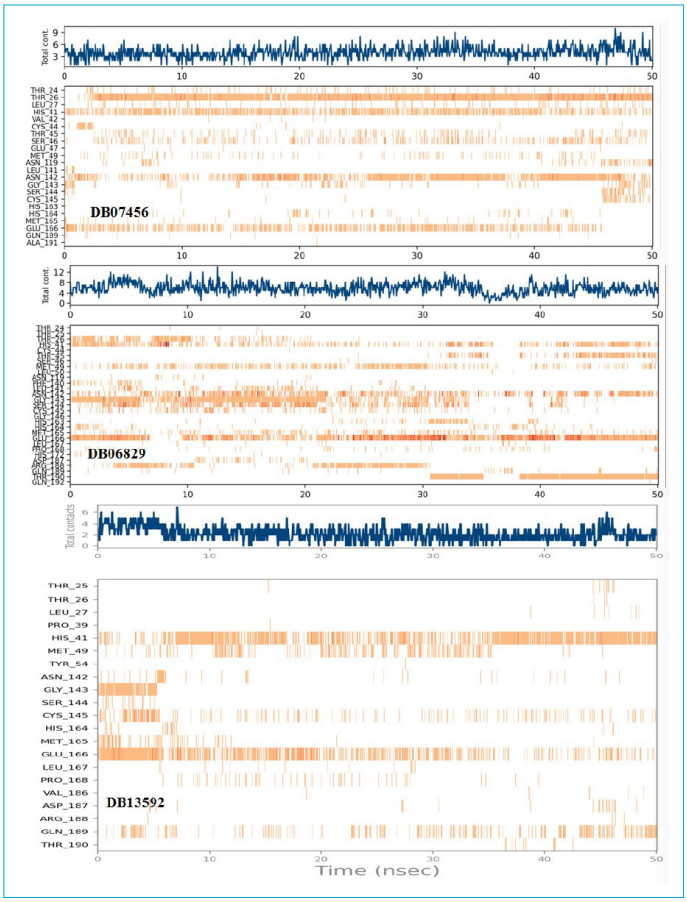

Figure 8. Timeline representation of interactions and contacts (H-bonds, hydrophobic, water bridges) in compound DB07456-[main protease (Mpro)] complex, DB06829-Mpro complex, and DB13592-Mpro complex during 50 ns molecular dynamics simulation.

combat the disease. In-silico study resulted in 6 hits with fitness to a common pharmacophore of SARS-CoV inhibitors, well-positioned in the pockets, and good binding affinity with the Mpro of SARS-CoV-2. The DrugBank identifiers of these hits are DB06829, DB07456, DB07458, DB13592, DB06834, DB11903, and DB03777. The binding energy values of DB07456, DB06834, and DB03777 were greater than that of remdesivir, and their corresponding Ki values were also favorable compared with remdesivir. The inhibitors of DB07456 showed good binding affinity with the Mpro predicted by the software. This molecule also interacted with the crucial amino acid residues HIS-41 and CYS-145. The 50 ns MD simulation study of the 3 best hits revealed DB07456 and DB13592 as potential SARS-CoV-2 Mpro inhibitors. These are commercially available and therefore, these hits may be useful to the scientific community for further in vitro and in vivo study to combat SARS-CoV-2. The VS studies performed thus far by different groups were based on the structure-based method only, but in this study, we employed pharmacophore-based VS as well as structure-based VS, and a MD simulation study, which may eliminate false-positive results.

Disclosures

Acknowledgements: The authors are grateful to the Schrödinger office of Bangalore, India for providing software for this research.

Peer-review: Externally peer-reviewed.

Conflict of Interest: None declared.

Authorship Contributions: Concept – S.D.N.; Design – B.D.N.; Supervision – S.D.N.; Data collection &/or processing – A.K.S., S.B.; Analysis and/or interpretation – B.D.N., S.B.; Literature search – S.B., A.K.S.; Writing – B.D.N., S.B.; Critical review – S.D.N., A.K.S.

References

1. Smith RD. Responding to global infectious disease outbreaks: lessons from SARS on the role of risk perception, communication and management. Soc Sci Med 2006;63:3113–23.
2. WHO. Middle East Respiratory Syndrome Coronavirus (MERS-CoV) – United Arab Emirates. Available at: https://www.who.int/csr/don/31-january-2020-mers-united-arab-emirates/en/. Accessed Jan 31, 2020.
3. Wu C, Liu Y, Yang Y, Zhang P, Zhong W, Wang Y, et al. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. Acta Pharmaceutica Sinica B 2020;10:766–88.
4. Pillaiyar T, Meenakshisundaram S, Manickam M. Recent discovery and development of inhibitors targeting coronaviruses. Drug Discov Today 2020;25:668–88.
5. Pillaiyar T, ManickamM, Namasivayam V, Hayashi Y, Jung SH. An Overview of Severe Acute Respiratory Syndrome–Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. J Med Chem 2016;59:6595–628.
6. Snijder EJ, Decroly E, Ziebuhr J. The Non-structural Proteins Directing Coronavirus RNA Synthesis and Processing. Adv Virus Res 2016;96:59–126.
7. Li X, Geng M, Peng Y, Meng L, Lu S. Molecular immune pathogenesis and diagnosis of COVID-19. J Pharm Anal 2020;10:102–8.
8. Dagur HS, Dhakar SS. Genome Organization of Covid-19 and Emerging Severe Acute Respiratory Syndrome Covid-19 Outbreak: A Pandemic. EJMO 2020;4:107–15.
9. Liu X, Wang XJ. Potential inhibitors for 2019-nCoV coronavirus M protease from clinically approved medicines. bioRxiv. 2020 Jan 29. doi: https://doi.org/10.1101/2020.01.29.924100. [Epub ahead of print].
10. Li G, Clercq ED. Therapeutic options for the 2019 novel coro-

navirus (2019-nCoV). Nat Rev Drug Discov 2020;19:149–50.
11. Contini A. Virtual Screening of an FDA Approved Drugs Database on Two COVID-19 Coronavirus Proteins. ChemRxiv. 2020 Feb 13. doi: https://doi.org/10.26434/chemrxiv.11847381.v1. [Epub ahead of print].
12. Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discov 2020;6:14.
13. Li Y, Zhang J, Wang N, Li H, Shi Y, Guo G, et al. Therapeutic Drugs Targeting 2019-nCoV Main Protease by High-Throughput Screening. bioRxiv. 2020 Jan 30. doi: https://doi.org/10.1101/2020.01.28.922922. [Epub ahead of print].
14. Narkhede RR, Cheke RS, Ambhore JP, Shinde SD. The Molecular Docking Study of Potential Drug Candidates Showing Anti-COVID-19 Activity by Exploring of Therapeutic Targets of SARS-CoV-2. EJMO 2020;4:185–95.
15. Tan YJ, Lim SG, Hong W. Characterization of viral proteins encoded by the SARS-coronavirus genome. Antiviral Research Antivir Res 2005;65:69–78.
16. Grahama RL, Sparks JS, Eckerle LD, Sims AC, Denison MR. SARS coronavirus replicase proteins in pathogenesis.Virus Res 2008;133:88–100.
17. Chen L, Gui C, Luo X, Yang Q, Günther S, Scandella E, et al. Cinanserin is an inhibitor of the 3C-like proteinase of severe acute respiratory syndrome coronavirus and strongly reduces virus replication in vitro. J Virol 2005;97:7095–103.
18. Jo S, Kim S, Shin DH. Inhibition of SARS-CoV 3CL protease by flavonoids. J Enzym Inhib Med Chem 2020;35:145–51.
19. Khaerunnisa S, Kurniawan H, Awaluddin R, Suhartati S, Soetjipto S. Potential Inhibitor of COVID-19 Main Protease (Mpro) from Several Medicinal Plant Compounds by Molecular Docking Study. Preprints. 2020 Mar 12. doi: 10.20944/preprints202003.0226.v1. [Epub ahead of prints].
20. Adem S, Eyupoglu V, Sarfraz I, Rasul A, Ali M. Identification of Potent COVID-19 Main Protease (Mpro) Inhibitors from Natural Polyphenols: An in Silico Strategy Unveils a Hope against CORONA. Preprints. 2020 Mar 21. doi: 10.20944/preprints202003.0333.v1. [Epub ahead of prints].
21. Liu X, Zhang B, Jin Z, Yang H, Rao Z. The crystal structure of COVID-19 main protease in complex with an inhibitor N3. Available at: https://www.wwpdb.org/pdb?id=pdb_00006lu7. Accessed Feb 5, 2020.
22. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 2018;46:D1074–D82.
23. Schrödinger, LLC, LigPrep, 2020. Available at: https://www.schrodinger.com/ligprep. Accessed Dec 4, 2020.
24. Schrödinger, LLC, Protein Preparation Wizard; Epik, 2020. Available at: https://www.schrodinger.com/protein-preparation-wizard. Accessed Dec 4, 2020.
25. Schrödinger, LLC, Prime, 2020. Available at: https://www.schrodinger.com/prime. Accessed Dec 4, 2020.
26. Jacobson MP, Pincus DL, Rapp CS, Day TJ, Honig B, Shaw DE, et al. A hierarchical approach to all-atom protein loop prediction. Proteins 2004;55:351–67.
27. Jacobson MP, Friesner RA, Xiang Z, Honig B. On the role of the crystal environment in determining protein side-chain conformations. J Mol Biol 2002;320:597–608.
28. Schrödinger, LLC, Phase, 2020. Available at: https://www.schrodinger.com/phase. Accessed Dec 4, 2020.
29. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, Friesner RA. PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. J Comput Aided Mol Des 2006;20:647–71.
30. Dixon SL, Smondyrev AM, Rao SN. PHASE: a novel approach to pharmacophore modeling and 3D database searching. Chem Biol Drug Des 2006;67:370–2.
31. Schrödinger, LLC, Glide, 2020. Available at: https://www.schrodinger.com/glide. Accessed Dec 4, 2020.
32. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 2004;47:1739–49.
33. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, et al. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem 2006;49:6177–96.
34. Schrödinger, LLC, QikProp, 2020. Available at: https://www.schrodinger.com/qikprop. Accessed Dec 4, 2020.
35. D. E. Shaw Research. Desmond Molecular Dynamics System. Available at: https://www.deshawresearch.com/resources_desmond.html. Accessed Dec 4, 2020.
36. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein M.L. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–35.
37. Yoshino R, Yasuo N, Sekijima M. Identification of key interactions between SARS-CoV-2 mainprotease and inhibitor drug candidates. Scientific Reports 2020;10:12493.
38. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell D, et al. Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity. J Comput Chem 2009;30:2785–91.
39. Mukherjee P, Shah F, Desai P, Avery M. Inhibitors of SARS-3CL-pro: virtual screening, biological evaluation, and molecular dynamics simulation studies. J Chem Inf Model 2011;51:1376–92.
40. Yang H, Yang M, Ding Y, Liu Y, Lou Z, Zhou Z, et al. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. Proc Natl Acad Sci USA 2003;100:13190–5.
41. Tu YF, Chien CS, Yarmishyn AA, Lin YY, Luo YH, Lin YT, et al. A Review of SARS-CoV-2 and the Ongoing Clinical Trials. Int J

Mol Sci 2020;21:2657.
42. Kufareva I, Abagyan R. Methods of protein structure comparison. Methods in Molecular Biology 2012;857:231–57.
43. Wang J. Fast Identification of Possible Drug Treatment of Coronavirus Disease-19 (COVID-19) through Computational Drug Repurposing Study. J Chem Inf Model 2020;60:3277–86.

Supplementary Materials

In silico identification of potential inhibitors of the main protease of severe acute respiratory syndrome coronavirus 2 (SRS-CoV-2) using a combined ligand-based and structure-based drug design approach.

Bimal Debnath,[1] Apu Kr. Saha,[2] Samhita Bhaumik,[3] Sudhan Debnath[4]

[1]Department of Forestry and Biodiversity, Tripura University, Suryamaninagar, Tripura, India
[2]Department of Mathematics, National Institute of Technology, Agartala, Tripura, India
[3]Department of Chemistry, Women's College, Agartala, Tripura, India
[4]Department of Chemistry, MBB College, Agartala, Tripura, India

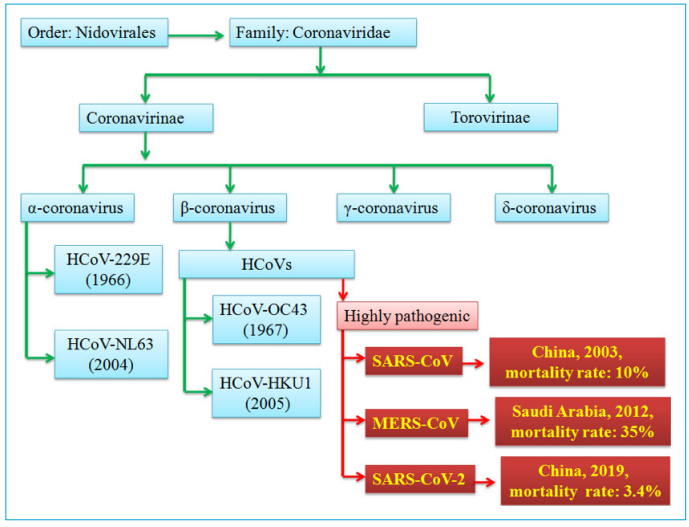

Figure S1. Classification of different human coronaviruses (HCoVs), severe acute respiratory syndrome (SARS-CoV), and Middle East Respiratory syndrome (MERS-CoV).[1′,2]

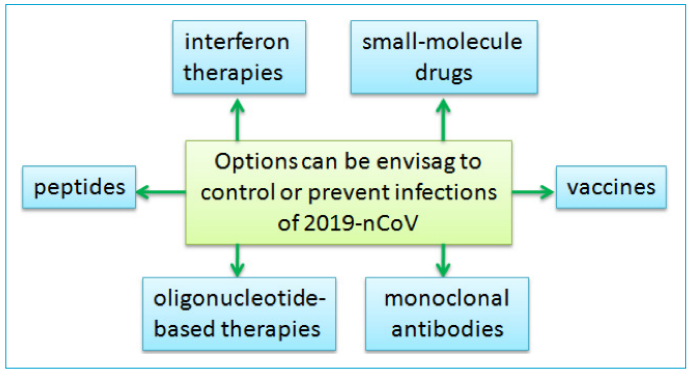

Figure S2. Different strategic options to prevent coronavirus 2019. [3]

References

1′. Xia S, Yan L, Xu W, Agrawal AS, Algaissi A, Tseng CK, et al. A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike. Sci Adv 2019;5:eaav4580.
2′. Tok TT, Tatar G. Structures and Functions of Coronavirus Proteins: Molecular Modeling of Viral Nucleoprotein. Int J Virol Infect Dis 2017;2:001-007.
3′. Li G, De Clercq E. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). Nat Rev Drug Discov 2020;19:149-50.
4′. Tan YJ, Lim SG, Hong W. Characterization of viral proteins encoded by the SARS-coronavirus genome. Antiviral Res 2005;65:69-78.

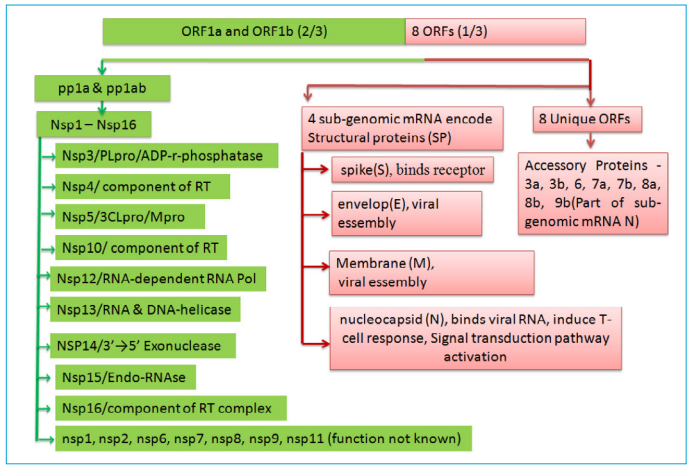

Figure S3. Summary of protein organization encoded by the severe acute respiratory syndrome genome.

3CLpro: Chymotrypsin-like protease; ADP-r-phosphatase: Adenosine diphosphate ribose-1-phosphatase; AMDRMT: S-adenosylmethionine-dependent ribose 2′-O-methyltransferase; Mpro: Main protease; Nsps: Nonstructural proteins; ORF: Open reading frame; pp1a and pp1ab: Polyproteins 1a & 1ab; PLpro: Papain-like cysteine protease; RdRP: RNA-dependent RNA-polymerase; RT: Replicase-transcriptase.[4′]

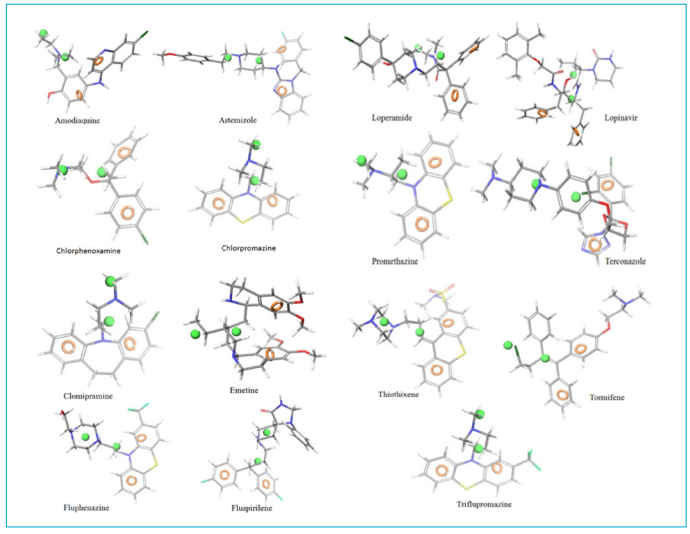

Figure S4. The 15 pharmacophore-matched severe acute respiratory syndrome (SARS-CoV) inhibitors.

Green: Hydrophobic; Ring aromatic: Orange.

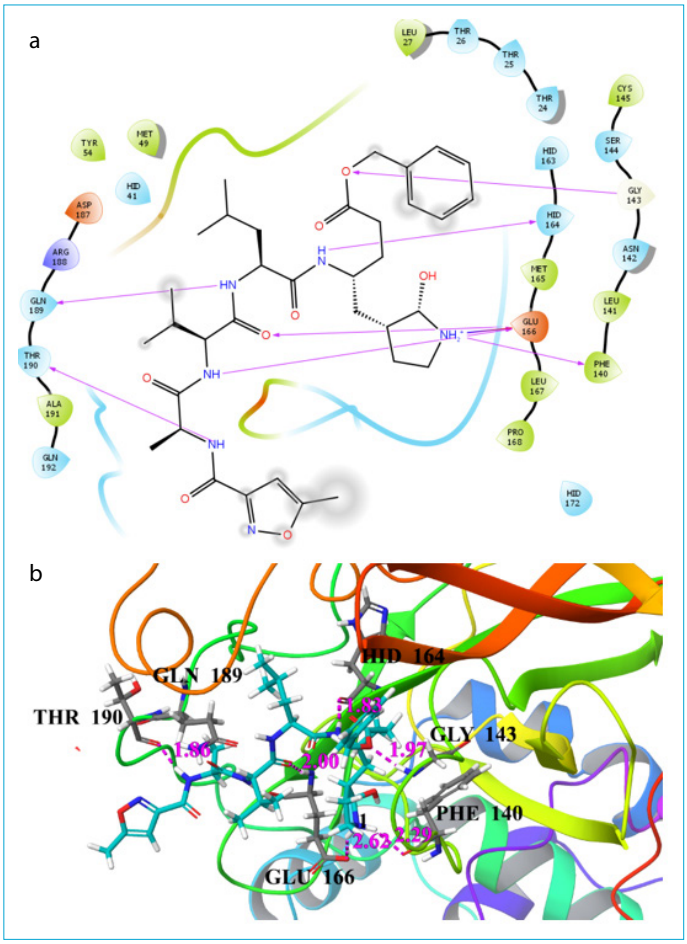

Figure S5. The original crystallographic-bound pose of carbon monoxide (CO) ligand (N3) of severe acute respiratory syndrome 2 (SARS-CoV-2) main protease protein (PDB ID: 6LU7). **(a)** 2D ligand interaction, **(b)** 3D interaction diagram. The CO ligand is light blue. The covalent bond with CYS-145 was broken before taking the image.

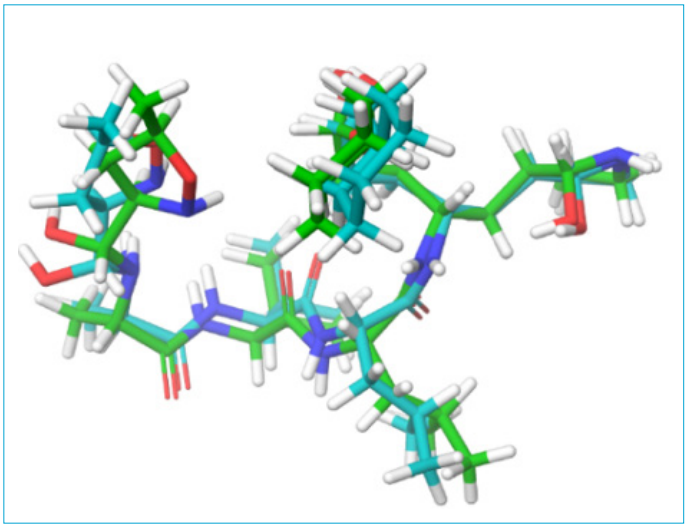

Figure S6. The best docked pose of the carbon monoxide (CO) ligand (light blue) superimposed on the original crystallographic-bound conformation (green).

Figure S7. The 3D ligand interaction diagram of remdesivir with severe acute respiratory syndrome 2 (SARS-CoV-2) main protease protein (PDB ID: 6LU7). Purple dotted line indicates H-bond donor-acceptor interactions and distance.

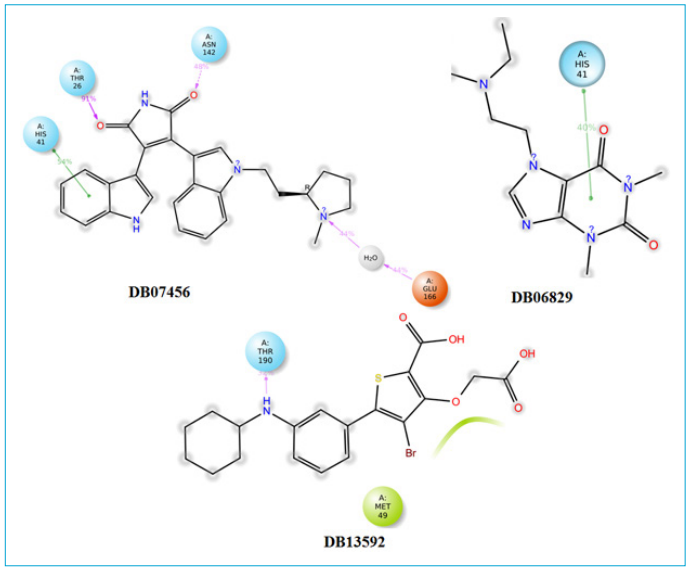

Figure S8. The contacts between DB07456, DB06829, and DB13592 and amino acid residues during molecular dynamics simulation.

Table S1. The existing targets of selected inhibitors

Compound ID	Target protein
DB06829	Tyrosine-protein phosphatase which acts as a regulator of endoplasmic reticulum unfolded protein response. Mediates dephosphorylation of EIF2AK3/PERK; inactivating the protein kinase activity of EI.
DB07456	Serine/threonine kinase which acts as a master kinase, phosphorylating and activating a subgroup of the AGC family of protein kinases. Its targets include: protein kinase B (PKB/AKT1, PKB/AKT2, PKB...
DB13592	Cytochromes P450 are a group of heme-thiolatemonooxygenases. In liver microsomes, this enzyme is involved in an NADPH-dependent electron transport pathway. It oxidizes a variety of structurally un...
DB06834	Non-receptor tyrosine kinase which mediates signal transduction downstream of a variety of transmembrane receptors including classical immunoreceptors like the B-cell receptor (BCR). Regulates seve...
DB11903	This drug entry is a stub and has not been fully annotated. It is scheduled to be annotated soon.
DB03777	Serine/threonine kinase which acts as a master kinase, phosphorylating and activating a subgroup of the AGC family of protein kinases. Its targets include: protein kinase B (PKB/AKT1, PKB/AKT2, PKB...